On principles, laws and theory in population ecology

A. A. Berryman, Dept of Entomology, Washington State Univ., Pullman, Washington 99164, USA (berryman@mail.wsu.edu).

Several commentaries have appeared recently on the status of ecology as a science, with some being quite pessimistic when comparing ecology to the hard sciences like physics. For example, some ecologists wonder if we will ever have general laws and theories (Roughgarden 1998, Lawton 1999) or become a predictive science (McIntosh 1985, Peters 1991). The reason for this, says Murray (1992), is that “biologists do not think like physicists”. Others think that ecology is just too variable and complicated to be subject to general theory (Hansson 2003). Some are more optimistic. For example, Ginzburg (1986) shows how ecologists can and do think like physicists and suggests that, by thinking this way, we can open new doors to ecological understanding (see also Colyvan and Ginzburg 2003Colyvan and Ginzburg 2003b, Ginzburg and Colyvan 2003). Berryman (1999) claims that population ecology already has a set of well-defined principles that are sufficient to describe, classify and explain all known patterns of population dynamics. Turchin (2001) and Colyvan and Ginzburg (2003a) are of a similar mind, suggesting that we may already have general laws resembling those of classical physics. In this paper I continue the philosophical debate from my particular point of view, and also attempt to clarify and consolidate differing views about ecology and its relation to the physical sciences.

The nature of principles, laws and theories

Before proceeding it is helpful to clearly define what we are talking about. Webster’s New World Dictionary (1958) defines a principle as “the ultimate source, origin or cause of something... a fundamental truth... an essential element, constituent, or quality, especially one that produces a specific effect”; a law of nature as “a sequence of events... that has been observed to occur with unvarying uniformity under the same set of conditions”, or “an exact formulation of the principle operating” in an observed regularity; a theory as “a formulation of apparent relationships or underlying principles of certain observed phenomena which has been verified to some degree”; and a theorem as “a proposition that is not self-evident but that can be proved from accepted premises and so is established as a law or principle”. Notice that theories and laws are often composed of, or can be derived from, one or more principles. In other words, principles define the fundamental components or ingredients of some larger body of knowledge, such as a theory. Principles may also be self-evident truths, or widely accepted premises, as in “first principles”. Laws, on the other hand, specify the rules that must be obeyed under given circumstances. They may be purely empirical regularities, in the sense that they are observed to occur with unvarying uniformity, but the reason why they do so may remain unknown. Laws may also be deduced, logically, from one or more basic propositions or principles as, for instance, theorems. It is important to realize that laws usually have conditional constraints, in the sense that they are obeyed under certain set of conditions. Thus, laws should not be expected to hold under all possible circumstances, something that is not always appreciated (Colyvan and Ginzburg 2003a). Sometimes principle and law are used synonymously, as in a basic principle that must be obeyed under certain conditions.

The nature of ecology

The science of ecology deals with interactions between living organisms and their environments. Thus, life sets ecology apart, at least to some extent, from the inanimate physical sciences. This fact may lead some to think that ecology should have its own special laws. On the other hand, ecology must surely be subject to the laws of the primary sciences (Colyvan and Ginzburg...
The first principle (geometric growth)

Ecologists seem to agree, in general, that geometric (exponential) growth is a good candidate for a law of population ecology (Ginzburg 1986, Turchin 2001). Hence the common epithet “Malthusian law”. However, since the geometric progression applies to many other kinds of systems, including compound-interest savings accounts and radioactive decay, one could question whether it should be considered a law of ecology at all. From this point of view it may be better to think of it as a general (mathematical) law to which biological populations also conform. On the other hand, since geometric growth is a fundamental and self-evident property of all populations living under a certain set of conditions (unlimited resources), I prefer to think of it as the first founding principle of population dynamics (Berryman 1999) or, if you prefer Malthus’ principle.

Ginzburg (1986) suggests an intriguing way of viewing the first principle by analogy with Newtonian mechanics. Newton’s first law (the law of inertia) states that all inanimate bodies move with uniform motion in a straight line unless affected by external forces. Similarly, we could think of geometric growth as an “ecological law of inertia”, since all populations of living organisms grow geometrically when unaffected by their environments. This is more than a mere curiosity since Newton’s first law completely revolutionized physics and forms the foundation for modern mechanics (Colyvan and Ginzburg 2003b).

Although my approach to the problem does not draw on analogies with Newtonian mechanics, it is somewhat similar to Ginzburg’s (Berryman 1999). The first principle can be explicitly stated as

\[
\frac{d}{dt} (\ln N) = R = \text{const}
\]  

where \( N \) is the size or density of the population and \( R \) is the per-capita logarithmic rate of change (I follow the convention of Royama (1977) in using \( R \) for the instantaneous (logarithmic) growth rate and reserving \( r \) for statistical correlation). In other words, the first principle states that all populations grow at a constant logarithmic rate unless affected by other forces in their environment. It is what Berryman and Turchin (2001) call the “null model” but its similarity to Newton’s first law is obvious. Given this fact, then it is immediately apparent that the problem of explaining and predicting the dynamics of any particular population boils down to defining how \( R \) deviates from the expectation of uniform growth (Berryman and Turchin 2001), or what I call the \( R \)-function

\[
R = f(B,G,P)
\]  

where \( B \) represents a set of biotic factors (populations of different species, including the one in question), \( G \) their genetic properties, and \( P \) a set of abiotic factors. This is basically a restatement of Lotka’s (1925) fundamental equations of kinetics, except that the point of reference is now the logarithmic per-capita rate of change (uniform growth) rather than the growth rate of the total population. This idea is quite similar to Ginzburg’s (and Newton’s first law) since we are able to ignore, or rather put aside, the obvious fact of uniform geometric growth and focus, instead, on the forces that affect the per-capita rate of change. This all makes sense because the forces of the environment actually affect the birth and death rates of individual organisms, as summarized by changes in \( R \).

At this point, however, Ginzburg (1986) diverges from my way of thinking. Instead of concentrating on the \( R \)-function, he continues the analogy with physics by focusing attention on the second derivative of the geometric progression

\[
\frac{d^2}{dt^2} (\ln N) = 0 \quad \text{or} \quad \frac{dR}{dt} = 0
\]  

In other words, he suggests that we should look at the forces affecting the rate of change of the per-capita growth rate, which is analogous to the idea of acceleration in mechanics. It is my position, however, that this is an unnecessary complication, and it is important for me to explain why. First, Newton’s law of inertia was essential for the development of theoretical physics because it allowed him to ignore the unknown forces that determined the initial velocity (uniform motion) of celestial bodies. By focusing on the second derivative (acceleration) he was able to deduce the local laws governing the motion of planets and other inanimate objects. In contrast, we know that uniform growth in biological populations is a manifestation of the general law of geometric progression. We also know that devia-
tions from uniform growth are determined, in a global sense, by forces that affect individual birth and death rates. This knowledge, in my opinion, makes it unnecessary to view population dynamics in the more difficult world of the second derivative, particularly since second order effects are dealt with under the fourth principle (see below).

My second point involves the fact that, unlike most physical systems, changes in populations of living organisms are often dependent, to a greater or lesser extent, on present or past states of the system. In other words, we know that changes in population density depend on \( R \) (under the first principle), but we also expect \( R \) to be affected by past population density; i.e. \( R = f(N_{t-d}) \), with \( d \) the delay in the response of \( R \) to changes in \( N \). Because of this, ecological systems are often subjected to feedback with a variable delay, and this has important consequences for their dynamics (Hutchinson 1948, May 1973, Berryman 1981, 1989).

From this perspective, I suppose, ecology can be viewed as a more complicated science than physics since feedback structures can give rise to extremely complex emergent dynamics (Milsum 1968). I should make clear at this point that I am not rejecting the second derivative as a valid way of thinking about ecological dynamics but, rather, that it may not be as necessary for the development of ecological theory as it was for physics. Finally, the concept of the \( R \)-function seems to form a concise and comprehensive framework for describing and explaining the dynamics of ecological systems in general, for analyzing and modeling ecological data, and for predicting ecological change (Royama 1977, 1992, Dennis and Taper 1994, Berryman 1999, Turchin 2001, Sibly and Hone 2002). It is not clear to me how a theory based on the second derivative can improve this situation.

The second principle (cooperation)

If we accept the first principle as the baseline for analyzing ecological change (and it is difficult to see how we can avoid this), then the crux of the problem is to identify and describe the forces that cause deviations from this baseline. In ecology, such deviations can only be brought about by environmental variables that affect the survival and reproduction of individual organisms, the basic components of \( R \) (at least in a global sense, where the redistribution of individuals in space does not contribute to population change). Some forces affect individual performance directly but are not themselves affected by the population. Others respond to the density of the population and so become involved in mutually causal (feedback) processes. Feedback is particularly important since it determines the stability properties of dynamic systems in accordance with general systems (or mathematical) principles (Milsum 1968, Berryman 1981, 1999). For this reason, the following discussion is largely concerned with the identification and definition of ecological feedback structures and their effects on population dynamics.

The second principle arises from the basic ecological premise that individuals can receive increasing benefits, in terms of higher reproduction and/or survival (or higher \( R \)), from increases in population density (Berryman 1999). For example, a higher probability of finding a mate, obtaining food (group hunting), or escaping enemies (group defense) – what is generally known as intra-specific cooperation or the Allee effect (Allee 1932). However, there must be an upper bound to this effect since all organisms have a maximum reproductive potential. These two premises lead to the formal proposition that, under the influence of the second principle, \( dR/dN > 0 \) and \( d^2R/dN^2 < 0 \). In other words, the \( R \)-function for a population operating under the second principle must rise at a decreasing rate with population density until it reaches some maximum level characteristic of a particular species and its physical environment (Fig. 1). Notice that the second principle can be stated as a theorem since it can be derived, logically, from two basic premises, the benefits of aggregation and finite reproduction.

There are a number of ways to express the second principle mathematically (Dennis 1989), but the one I prefer is (because it conforms with later equations)

\[
R = A - B \frac{1}{N_{t-d}} = A \left[ 1 - \left( \frac{E}{N_{t-d}} \right)^U \right] \tag{4}
\]

where \( A \) is the maximum per-capita rate of change of the species in a particular environment, \( B \) is a constant of proportionality, \( E \) is an equilibrium point, \( N_{t-1} \) is the density of the population \( d \) units of time in the past, and \( U \) is a coefficient that allows for non-linear re-

---

Fig. 1. Hypothetical \( R \)-function for a population under the influence of the second principle: \( R \) = realized per-capita rate of change, \( N \) = population density, \( A \) = maximum per-capita rate of change in a given environment, \( E \) = unstable threshold or escape point. \( R \)-function calculated from the equation \( R = A - B/N^U \), with \( A = 1 \), \( B = 50 \) and \( U = 1 \).
responses to density. Notice that the equilibrium point, E, occurs where the R-function crosses the \( R = 0 \) axis (Fig. 1), and that it is unstable since the population grows when \( N > E \) (because \( R > 0 \)) and declines towards extinction when \( N < E \) (because \( R < 0 \)). For this reason, \( E \) is sometimes called an extinction threshold.

The type of feedback (+ or −) acting on a population (other than geometric growth, which itself causes + feedback) is given by the derivative (or slope) of the R-function. In the case of the second principle \( dR/dN > 0 \), so the feedback is positive. From a general system’s viewpoint we know that positive feedback causes deviations to be amplified with time and, for this reason, we expect population trajectories to accelerate away from the reference trajectory (uniform growth) when under the influence of the second principle. This is sometimes called hyper-geometric growth (or decline).

As can be seen in Fig. 1, positive feedback can also create unstable thresholds (boundaries) separating qualitatively different kinds of dynamic behavior; e.g. growth from extinction, pest outbreaks from insignificance, collapses of natural resources from previous abundance. Hence, the second principle explains the occurrence of metastable dynamics and multiple stable states in ecological systems (Holling 1973, May 1977, Berryman 1978, 1999). Despite this, the second principle rarely receives much attention in the ecological literature (but see Dennis 1989, Courchamp et al. 1999, Stephens and Sutherland 1999). Popular textbooks often ignore Allee’s principle completely (Ricklefs 1990a, b), or give it short shrift (Begon et al. 1990). Turchin (2001) does not list it amongst his “laws” of population dynamics, nor does he consider it an important component of “complex population dynamics” (Turchin 2003), even though the second principle gives rise to the most complex population dynamics of all (multiple stable states). Royama (1992) has little to say about Allee effects in his book on “analytical population dynamics”, and seems to reject the idea of metastability in ecology altogether. It seems strange that Allee’s principle is so widely neglected by ecologists when it is so obviously essential to any general theory of population dynamics, for without it we cannot explain the dynamics of extinction and multiple coincident stable states.

The third principle (competition)

The third principle arises from the basic ecological premise that organisms have problems acquiring resources, or become more vulnerable to natural enemies, as their populations become larger, and this results in lowered reproduction and/or survival; i.e. the derivative of the \( R \)-function is \( dR/dN < 0 \). This phenomenon is usually called intra-specific competition for limited resources, including the idea of “enemy free space” as a resource (Jeffries and Lawton 1984). In fact, competition can often be reduced to a problem of insufficient space – space in which to gather resources or to hide from or escape enemies (Berryman 1999). Thus, we expect the \( R \)-function for a population subjected to the third principle to decline continuously with population density (Fig. 2). One way to express this is

\[
R = A - BNQ, \quad \text{with} \quad A = 1, \quad B = 0.07, \quad \text{and} \quad Q = 0.5
\]

where \( A \) is, once again, the maximum per-capita rate of change in a given environment, \( B \) is a constant of proportionality, \( K \) is the equilibrium density, and \( Q \) is a coefficient that allows for the nonlinear density effects so often observed in nature (Richards 1959, Nelder 1961). This particular \( R \)-function is a generalization of the famous “logistic” equation (Verhulst 1838, Lotka 1925), which may lead us to call it Verhulst’s principle. Notice that the equilibrium, \( K \), is stabilizing since the population grows when \( N < K \) (because \( R > 0 \)) and declines when \( N > K \) (because \( R < 0 \)), resulting in a trajectory that tends towards \( K \). Hence, the principle of competition acts as a stabilizing influence on population dynamics. Notice also that the third principle defines a negative feedback loop since \( dR/dN < 0 \) for all \( N \) and, as a result, acts in opposition to the first and second principles. In other words, competitive interactions cause the rate of population growth to decelerate towards zero, in a similar way to friction acting on the motion of a physical body (Ginzburg 1986, Colyvan and Ginzburg 2003b). Finally, I should note that Turchin (2003) and some others use the term “self-limitation” instead of “competition” for the mechanism underlying the third principle. In my opinion, however, competition is the more appropriate word since it has
The fourth principle (interacting species)

The ecological principles discussed so far result from the effects of individuals of the same species upon each other. In other words, they involve intra-specific effects. In contrast, the next two principles address the fact that natural populations are embedded within a web of interactions with other organisms, as well as their physical environment, and this can give rise to feedback loops involving more than one dynamic agent. For example, inter-specific interactions between predators and their prey can create negative feedback between the two species, in the sense that increases in prey numbers result in higher predator numbers (through increased reproduction) and this feeds back to reduce prey numbers (through increased mortality). Consider the general R-functions for prey and predator populations

\[
R^N = f^N(N_{t-1}, P_{t-1})
\]

\[
R^P = f^P(P_{t-1}, N_{t-1})
\]

where \(R^N\) and \(R^P\) are the per-capita rates of change of prey and predator, respectively, and \(f^N\) and \(f^P\) are unspecified functions of prey and predator densities, \(N_{t-1}\) and \(P_{t-1}\), at the beginning of a period of time. Royama (1977) shows how such a system of two first order equations can be reduced to a second order equation for one species; i.e. for the prey we would get

\[
R = f(N_{t-1}, N_{t-2})
\]

Notice that the per-capita rate of growth of the prey population now depends on its density one and two time steps previously. In other words, the prey R-function is controlled by both first and second order feedback. In more general terms, (Eq. 7) can be written for an unspecified maximum time delay \(d\)

\[
R = f(N_{t-1}, N_{t-2}, \ldots, N_{t-d})
\]

Assuming that \(f(\ )\) can be approximated by a linear function, we obtain the explicit relationship

\[
R = a_0 + a_1 N_{t-1} + a_2 N_{t-2} + \cdots + a_d N_{t-d}
\]

frequently used in ecological time series analysis (Berryman 1999, 2002). For statistical reasons, population densities on the right hand side of (9) are often transformed to logarithms prior to statistical analysis (Royama 1992, Berryman and Turchin 2001), or more complex models may be used to improve the fit (Turchin and Taylor 1992, Bjornstad et al. 1995).

As we have seen in Eq. 7, mutual interactions between populations and components of their environments (e.g. predators) can produce delays in the feedback between them. Furthermore, we know from general systems theory that time-delays can generate oscillatory instability in the variables involved in negative feedback loop. This leads to the proposition that oscillatory (cyclical) dynamics are likely to be seen when populations are involved in negative feedback with other species, or even physical components of their environments (Hutchinson 1948, Royama 1977, Berryman 1981, 1999, 2002). Because consumer-resource interactions create the conditions for negative feedback and are also ubiquitous components of ecological systems, Turchin (2001) calls this “the law of consumer-resource oscillations”. However, this principle obviously has much more general roots, and can also apply to negative feedback between populations and their physical environment (e.g. global warming) or even to intrinsic mechanisms like maternal effects (Ginzburg and Taneyhill 1994). For these reasons I prefer to think of it as a general principle underlying and explaining cyclic population oscillations (Berryman 1999, 2002).

The fifth principle (limiting factors)

The fifth principle also considers the fact that populations can be affected by many feedback loops involving many different species and physical factors. If all these potential feedback processes were to operate simultaneously, then the dynamics would usually be extremely complex, or even chaotic. The fact is, most natural populations are characterized by first or second order dynamics, a few by third order, and none as far as I know by higher order dynamics. This suggests that only one or two other species dominate the feedback structure of a population at any one time and place (Berryman 1993). These are often referred to as limiting factors, an idea that dates back to Liebig’s (1840) “law of the minimum”, and is supported by Paine’s (1980, 1992) experiments showing that “strong interactions” dominate the dynamics of inter-tidal food webs. Although there is some controversy over the details of this principle (e.g. what constitutes a limiting factor), the fact remains that natural populations rarely exhibit high order (e.g. chaotic) dynamics, and this implies that they are controlled by simple feedback architectures.

The fifth principle recognizes that the control of population dynamics can change from time to time and place to place as feedback processes change in response to environmental conditions and/or population density. Feedback mechanisms that change with respect to population density are particularly important since they can affect the shape of the R-function. For example, R-functions for the spruce budworm (Ludwig et al. 1978) and gypsy moth (Campbell and Sloan 1978) may
have three equilibrium points, two stable and one unstable (Fig. 3), and can exhibit complex (metastable) dynamics in variable environments.

Conclusions

Turchin (2001) asks “Does population ecology have general laws?” Certainly we have a set of basic (foundingational) principles that, as far as I can see, are sufficient to describe, classify, explain, and predict (at least in a qualitative sense) the dynamics of any and all populations of living organisms (Berryman 1999). In other words, some of us think that we already have a grand explanatory theory (similar to the theory of evolution), although we may disagree somewhat over the importance of the basic principles and how they should be formulated (Ginzburg and Colyvan 2003a, Turchin 2001). Most ecologists would probably agree that a comprehensive theory of population dynamics should include, at the very least, the first four principles enumerated above. Suppose we can all agree on the theory, then what does this mean for the laws of ecology? For example, should we upgrade the five principles to laws? We already have Malthusian law and Liebig’s law for the first and fifth principles. Few would argue, I suspect, with the former but there may be some debate over the latter. What about Alle’s law for the second principle, Verhulst’s law for the third, and Hutchinson’s law for the fourth? A good case could probably be made for any or all of these. In fact, both Pearl (1924) and Lotka (1925) considered the elementary logistic equation (a simple model for the third principle) as the “law of population growth”. Many ecologists disagree, however, citing its failure to describe the growth of many real populations (Fagerström 1987, Peters 1991, Turchin 2001). On the other hand, if we accept the definitions cited above, and if we agree with Colyvan and Ginzburg (2003a) that laws are not meant to be infallible, then the logistic looks like a good candidate since it accurately describes the growth of populations growing under a particular (idealized) set of conditions (Gause 1934, Allee et al. 1949).

Colyvan and Ginzburg (2003a) also suggest that the laws of nature should define the places in the grand theory where explanation ends, or is unnecessary. In the case of the five principles of population dynamics, explanation ends with general systems theory. In other words, we accept the general propositions that positive feedback produces deviation amplification, negative feedback deviation attenuation, time delays oscillatory instability, and so on. Are these the laws we seek? If so, they are not peculiar to ecology. Perhaps this is why population ecology does not have, and does not need, its own laws. As I remarked earlier, ecology is an integrative science that must be subject to the more general rules of complex integrative systems, as well as the more basic laws of physics and chemistry. Does this make ecology a less rigorous science? I, for one, do not see why it should. I think population ecology already has a strong, integrative theory that rests on five sound ecological principles. So what if the underlying laws do not belong exclusively to ecology or the life sciences? I do not think this is reason to be ashamed of our science, or to denigrate it in comparison to the physical sciences? On the other hand, I do wish ecologists would desist from inventing new terminology, for it gives the impression that we are continuously searching for new theories and implies, incorrectly in my opinion, that we have none. Rather than inventing new terms, I think we should use those of the more basic or integrative disciplines. Inventing new terms suggests that we are ignorant of these other disciplines, generates confusion, and gives an impression of immaturity and insecurity that I do not think we deserve (Berryman et al. 2002).

Acknowledgements – I would like to thank Lev Ginzburg, Mark Colyvan, Peter Turchin, Mauricio Lima, and Bradford Hawkins for their always stimulating and thoughtful discussions and for reading and commenting on earlier drafts of this paper.

References


Gause, G. F. 1934. The struggle for existence. – Williams and Wilkins.


Hansson, L. 2003. Why ecology fails at application: should we consider variability more than regularity? – Oikos 100: 624–627.


Liebig, J. 1840. Chemistry and its application to agriculture and physiology. – Taylor and Walton.


McIntosh, R. P. 1985. The background of ecology: concept and theory. – Cambridge Univ. Press.


